Amendments to the Specification:

At page 7, lines 7-29, please substitute the following paragraph for the existing text:

In preferred embodiments of this method more than one SAGE tags are simulateously identified. The multiple identification provides fr high-throughput. The high-throughput generation of longer SAGE tags for gene identification (GLGI) prodecure has several important features, for example, (i) 3' cDNAs instead of full-length cDNAs are used as the templates for GLGI amplification. This prevents artificial amplification from non-specific annealing of snese primer. The 3' cDNAs can be amplified to provide sufficient templates for GLGI amplification; (ii) a single antisense primer (in one example the primer is: 5'-ACTATCTAGAGCGGCCGCTT-3' <u>(SEQ ID NO: 12)</u> (see also Example 3)) is used for all GLGI reactions instead of using combination of the five anchored oligo dT primers. The sequence of the antisense primer is located in 3′ end of all the cDNA templates incorporated from anchored oligo dT primers used for the first strand cDNA synthesis. Use of a single primer also increases the efficiency of GLGI amplification significantly as any annealing of this primer with 3' end sequence results in extension during PCR. This feature is particularly useful to amplify the templates with low copies; (iii) Use of PLATINUM Taq polymerase instead of Pfu DNA polymerase increases the yield of final products, while maintaining high specificity; (iv) the GLGI amplified DNAs are directly precipitated and cloned into vector without gel purification, which further prevents loss of amplified products. The inventors contemplate that this is especially important for products with short sizes and for products generated from templates with low copies. Thus, the methods of this invention provide the ability for large-scale identification of expressed genes. Genes of any eukaryotic origin, including human genes may therefore be identified at an accelerated rate by the simple, efficient and low-cost methods set forth herein.

At page 20 through page 24, please replace Table 1 with the following substitute table:

Table 1: Restriction Enzymes

Enzyme Name	Recognition Sequence	
AatII	GACGTC	
Acc65 I	GGTACC	
Acc I	GTMKAC	
Aci I	CCGC	
Acl I	AACGTT	
Afe I	AGCGCT	
Afl II	CTTAAG	
Afl III	ACRYGT	
Age I	ACCGGT	
Ahd I	GACNNNNNGTC	(SEQ ID NO: 36)
Alu I	AGCT	
Alw I	GGATC	
AlwN I	CAGNNNCTG	
Apa I	GGGCCC	
ApaL I	GTGCAC	
Apo I	RAATTY	
Asc I	GGCGCGCC	
Ase I	ATTAAT	
Ava I	CYCGRG	
Ava II	GGWCC	
Avr II	CCTAGG	
Bae I	NACNNNNGTAPyCN	(SEQ ID NO: 37)
BamH I	GGATCC	
Ban I	GGYRCC	
Ban II	GRGCYC	
Bbs I	GAAGAC	
Bbv I	GCAGC	
BbvC I	CCTCAGC	
Bcg I	CGANNNNNNTGC	(SEQ ID NO: 38)
BciV I	GTATCC	
Bcl I	TGATCA	
Bfa I	CTAG	
Bgl I	GCCNNNNNGGC	(SEQ ID NO: 39)
Bgl II	AGATCT	
Blp I	GCTNAGC	
Bmr I	ACTGGG	
Bpm I	CTGGAG	
BsaA I	YACGTR	
BsaB I	GATNNNNATC	(SEQ ID NO: 40)
BsaH I	GRCGYC	

Bsa I	GGTCTC	
BsaJ I	CCNNGG	
BsaW I	WCCGGW	
BseR I	GAGGAG	
Bsg I	GTGCAG	
BsiE I	CGRYCG	
BsiHKA I	GWGCWC	
BsiW I	CGTACG	
Bsl I	CCNNNNNNNGG	(SEQ ID NO: 41)
BsmA I	GTCTC	
BsmB I	CGTCTC	
BsmF I	GGGAC	
Bsm I	GAATGC	
BsoB I	CYCGRG	
Bsp1286 I	GDGCHC	
BspD I	ATCGAT	
BspE I	TCCGGA	
BspH I	TCATGA	
BspM I	ACCTGC	
BsrB I	CCGCTC	
BsrD I	GCAATG	
BsrF I	RCCGGY	
BsrG I	TGTACA	
Bsr I	ACTGG	
BssH II	GCGCGC	
BssK I	CCNGG	
Bst4C I	ACNGT	
BssS I	CACGAG	
BstAP I	GCANNNNNTGC	(SEQ ID NO: 42)
BstB I	TTCGAA	
BstE II	GGTNACC	
BstF5 I	GGATGNN	
BstN I	CCWGG	
BstU I	CGCG	
BstX I	CCANNNNNNTGG	(SEQ ID NO: 43)
BstY I	RGATCY	
BstZ17 I	GTATAC	
Bsu36 I	CCTNAGG	
Btg I	CCPuPyGG	
Btr I	CACGTG	
Cac8 I	GCNNGC	
Cla I	ATCGAT	
Dde I	CTNAG	
Dpn I	GATC	
Dpn II	GATC	
Dra I	TTTAAA	
Dra III	CACNNNGTG	
Drd I	GACNNNNNNGTC	(SEQ ID NO: 44)

Eae I	YGGCCR	
Eag I	CGGCCG	
Ear I	CTCTTC	
Eci I	GGCGGA	
EcoN I	CCTNNNNNAGG	(SEQ ID NO: 45)
EcoO109 I	RGGNCCY	
EcoR I	GAATTC	
EcoR V	GATATC	
Fau I	CCCGCNNNN	
Fnu4H I	GCNGC	
Fok I	GGATG	
Fse I	GGCCGGCC	
Fsp I	TGCGCA	
Hae II	RGCGCY	
Hae III	GGCC	
Hga I	GACGC	
Hha I	GCGC	
Hinc II	GTYRAC	
Hind III	AAGCTT	
Hinf I	GANTC	
HinP1 I	GCGC	
Hpa I	GTTAAC	
Hpa II	CCGG	
Hph I	GGTGA	
Kas I	GGCGCC	
Kpn I	GGTACC	
Mbo I	GATC	
Mbo II	GAAGA	
Mfe I	CAATTG	
Mlu I	ACGCGT	
Mly I	GAGTCNNNNN	(SEQ ID NO: 46)
Mnl I	CCTC	
Msc I	TGGCCA	
Mse I	TTAA	
Msl I	CAYNNNNRTG	(SEQ ID NO: 47)
MspA1 I	CMGCKG	
Msp I	CCGG	
Mwo I	GCNNNNNNNGC	(SEQ ID NO: 48)
Nae I	GCCGGC	
Nar I	GGCGCC	
Nci I	CCSGG	
Nco I	CCATGG	
Nde I	CATATG	
NgoMI V	GCCGGC	
Nhe I	GCTAGC	
Nla III	CATG	
Nla IV	GGNNCC	
Not I	GCGGCCGC	

Nru I	TCGCGA	
Nsi I	ATGCAT	
Nsp I	RCATGY	
Pac I	TTAATTAA	
PaeR7 I	CTCGAG	
Pci I	ACATGT	
PflF I	GACNNNGTC	
PflM I	CCANNNNNTGG	(SEQ ID NO: 49)
PleI	GAGTC	
Pme I	GTTTAAAC	
Pml I	CACGTG	
PpuM I	RGGWCCY	
PshA I	GACNNNNGTC	(SEQ ID NO: 50)
Psi I	TTATAA	
PspG I	CCWGG	
PspOM I	GGGCCC	
Pst I	CTGCAG	
Pvu I	CGATCG	
Pvu II	CAGCTG	
Rsa I	GTAC	
Rsr II	CGGWCCG	
Sac I	GAGCTC	
Sac II	CCGCGG	
Sal I	GTCGAC	
Sap I	GCTCTTC	
Sau3A I	GATC	
Sau96 I	GGNCC	
Sbf I	CCTGCAGG	
Sca I	AGTACT	
ScrF I	CCNGG	
SexA I	ACCWGGT	
SfaN I	GCATC	
Sfc I	CTRYAG	
Sfi I	GGCCNNNNNGGCC	(SEQ ID NO: 51)
Sfo I	GGCGCC	
SgrA I	CRCCGGYG	
Sma I	CCCGGG	
Sml I	CTYRAG	
SnaB I	TACGTA	
Spe I	ACTAGT	
Sph I	GCATGC	
Ssp I	AATATT	
Stu I	AGGCCT	
Sty I	CCWWGG	
Swa I	ATTTAAAT	
Taq I	TCGA	
Tfi I	GAWTC	
Tli I	CTCGAG	

Tse I	GCWGC	
Tsp45 I	GTSAC	
Tsp509 I	AATT	
TspR I	CAGTG	
Tth111 I	GACNNNGTC	
Xba I	TCTAGA	
Xcm I	CCANNNNNNNNNTGG	(SEQ ID NO: 52)
Xho I	CTCGAG	
Xma I	CCCGGG	
Xmn I	GAANNNNTTC	(SEQ ID NO: 53)

At page 41, lines 14-20, please substitute the following paragraph for the existing text:

SAGE Tags. A group of SAGE tags 10 bases long were selected from the SAGE tag sequences database generated from epithelium cells of normal colon (Zhang *et al.*, 1997) ~~(http://www.ncbi.nlm.nih.gov/SAGE/sagerec.cgi?rec=166)~~. Each selected SAGE tag sequence was searched in the UniGene database ~~(http://www.ncbi.nlm.nih.gov/SAGE/SAGEtag.cgi?tag)~~ to identify it as a matched or an unmatched tag sequence. Each matched sequence was given the appropriate Unigene ID number. Both matched and unmatched tags were used in the experiments.

At page 42, lines 16-17, please substitute the following paragraph for the existing text:

Database search. All the sequences generated from the clones were searched using the BLAST program for alignment ~~(http://www.ncbi.nlm.nih.gov/BLAST/)~~.

At page 51, lines 1-6, , please substitute the following recitation for the existing text:

All collected sequences were matched to GenBank Database (NR and ESTs, ~~http://www.ncbi.nlm.nih.gov/BLAST/)~~ through BLAST. Any mismatch between the SAGE tag sequence used for GLGI amplification and the SAGE tag sequence of the matched sequence in database was considered as non-specific amplification, and these sequences were eliminated from further analysis. The matched sequence ID was used to search UniGene database to obtain the UniGene cluster ID.

At page 44, line 26, to page 45, line17, please substitute the following recitation for the existing text:

Identify the correct sequence from multiple sequences that matched with the same SAGE Tag. When matching SAGE tag sequences in databases, a single SAGE tag may align with several sequences. For example, nine out of 40 SAGE tag sequences show matches to multiple Unigene Clusters (Zhang *et al.*, 1997). Other than sharing the same SAGE tag sequence, these matched sequences have no homology and are derived from various different tissues. To test this issue experimentally, 12 SAGE tags were used for amplification with cDNA samples from 24 different human tissues. Four out of these 12 tags generated multiple templates. For example, the SAGE tag (GTCATCACCA) (SEQ ID NO: 17) generated five different sequences from five different tissues (fetal liver, skeletal muscle, spinal cord, trachea and colon), and two different sequences from the same tissue (spinal cord) (Table 4). All of these fragments contained the same SAGE tag sequence, but the rest of the sequences showed no homology. Among these sequences, the ones from colon tissue all matched the previous amplified sequences in the colon (Table 3). These data indicate that a SAGE tag itself may not be sufficient to serve as a unique identifier for a particular sequence, when several sequences share the same SAGE tag sequences. It is important to distinguish which one of the matched sequences is the correct sequence corresponding to the particular SAGE tag. To avoid the uncertainty when different sequences are expressed from different tissues, it will be necessary to generate the fragment from the same tissue used to generate the SAGE tag. The inventors' observations also indicate that relying only on a database search to identify the sequence corresponding to a SAGE tag may provide misleading information. Direct amplification of the specific template with the inventors strategy will be very useful for confirmation of the validity of a particular SAGE tag.

At page 46, lines 1-5, please substitute the following recitation for the existing text:

Table 3. Summary of GLGI results from SAGE Tags

SAGE Tags (10 base)	Unigene ID	3' end nucleotide in matched sequences*	Amplified by anchored oligo dT	Length of sequence (bs)	Match to original sequence**
GGAAGGTTTA (SEQ ID NO: 18)	Hs.105484	dT/dG	dT	77	+
AGATCCCAAG (SEQ ID NO: 19)	Hs.50813	dC/dG	dC	84	+
CTTATGGTCC (SEQ ID NO: 20)	Hs.179608	dT	dT	86	+
AGGATGGTCC (SEQ ID NO: 21)	Hs.71779	dC	dC	112	+
GTCATCACCA (SEQ ID NO: 22)	Hs.32966	dC	dC	119	+
GACCAGTGGC (SEQ ID NO: 23)	Hs.143131	dC/dT	dC	135	+
CTGTTGGTGA (SEQ ID NO: 24)	Hs.3463	dC	dC	148	+
ACTGGGTCTA (SEQ ID NO: 25)	Hs.227823	dG	dG	150	+
TACGGTGTGG (SEQ ID NO: 26)	Hs.105460	dC	dC	166	+
CGGTGGGACC (SEQ ID NO: 27)	Hs.99175	dC/dT/dG	dC	200	+
CCTTCAAATC (SEQ ID NO: 28)	Hs.23118	dC/dT	dC	220	+
GGAGGCGCTC (SEQ ID NO: 29)	Hs.33455	dT/dG	dT	238	+
AAGAAGATAG (SEQ ID NO: 30)	Hs.73848	dT	dT	317	+
GATCCCAACT (SEQ ID NO: 31)	Hs.118786	dG/dT/dC	dG	329	+
GAACAGCTCA (SEQ ID NO: 32)	Hs.194659	dT	dG	382	+
AGGTGACTGG (SEQ ID NO: 33)	-	-	dC	156	-
CACCTAGTTG (SEQ ID NO: 34)	-	-	dT	170	-
CCTGTCTGCC (SEQ ID NO: 35)	-	-	dT	249	-

*The 3' end nucleotides from all the sequences were included in each matched Unigene cluster.
**The amplified sequences were matched to databases again. The last three sequences have no matches and represent novel sequences.

At page 47, lines 1-2, please substitute the following recitation for the existing text:

Table 4. Detection of heterogeneous sequences in various tissues containing the same SAGE Tag

SAGE TAG	Positive tissues	Unigene ID	length of sequence
CGGTGGGACC	Colon, Thymus, Small intestine	Hs.99175	200
(SEQ ID NO: 14)	Small intestine	no match	368
	Thymus	no match	90
AGATCCCAAG	Colon, Heart, Placenta, Thymus	Hs.50813	84
(SEQ ID NO: 15)	Placenta	no match	53
	Skeletal muscle	Hs.85937	282
	Testis	no match	227
	Thymus, Placenta	no match	51
CTTATGGTCC	Bone marrow	Hs.237416	393
(SEQ ID NO: 16)	Bone marrow	no match	144
	Colon	Hs.179608	86
GTCATCACCA	Fetal liver, Spinal cord	Hs.222346	125
(SEQ ID NO: 17)	Skeletal muscle	Hs.1288	399
	Spinal cord	Hs.9641	394
	Trachea	no match	225
	colon	Hs.32966	136

At page 48, lines 16-30, please substitute the following recitation for the existing text:

The same RNA samples from human and mouse myeloid cells used for SAGE analysis were used as the templates for GLGI amplification. mRNAs from 5 μg of total RNA of each sample were isolated with Oligo $(dT)_{25}$ Dynabeads (Dynal), following the manufacturer's protocol. Poly(dA/dT) cDNAs were synthesized using a cDNA synthesis kit (Cat. No: 18267-021, Life Technologies) and the 5' biotinylated, 3' anchored oligo (dT) primers were used for first strand cDNA synthesis (5' biotin-ATCTAGAGCGGCCGC-T16-A,G, CA,CG and CC) (SEQ ID NOS: 1-5) (Wang *et al.*, 2000). The double-strand cDNAs were then digested with *NlaIII*, and 3' cDNAs were isolated with streptavidin beads (Dynal), following the manufactures protocol. In order to generate enough 3' cDNAs for GLGI analysis, 3' cDNA templates were amplified by PCR as the following: SAGE linker A or B was ligated to the 3' cDNAs bound to the beads (Linker A: 5′-TTTGGATTTGCTGGTGCAGTACAACTAGGCTTAATAGGGACATG - 3′ (SEQ ID NO: 6)

and 5′- $_{P}$TCCCTATTAAGCCTAGTTGTACTGCACCAGCAAATCC [amino mod. C7]- 3′ (SEQ ID NO: 7); or Linker B: 5′- TTTCTGCTCGAATTCAAGCTTCTAACGATGTACGGGGA CATG - 3′ (SEQ ID NO: 8) and 5′- $_{P}$TCCCCGTACATCGTTAGAAGCTTGAATTCGAGCAG [amino mod. C7]- 3′) (SEQ ID NO: 9) ~~(http://www.sagenet.org/sage_protocol.htm)~~. The ligated 3' cDNAs were then amplified by 20 cycles of PCR at 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s, with PLATINUM Taq polymerase (Life Technologies), SAGE sense primer (5′- GGATTTGCTGGTGCAG TACA - 3′ (SEQ ID NO: 10) for linker A; or 5′- CTGCTCGAATTCAAGCTTCT - 3′ (SEQ ID NO: 11) for linker B)~~(http://www.sagenet.org/sage_protocol.htm)~~ and antisense primer (5′ - ACTATCTAGAGCGGCCGCTT- 3′) (SEQ ID NO: 12) located in the 5′ end of anchored oligo dT primers used for the first strand cDNA synthesis. The amplified templates were extracted by phenol/chloroform, precipitated by ethanol/NH_4OAc/glycogen, and resuspended in TE buffer for GLGI amplification.

At page 49, line 12, to page, line 29, please substitute the following recitation for the existing text:

The sense primer used for GLGI amplification included 14 bases (CATG + 10 base SAGE tag sequence) at the 3′ end and 6 bases (GGATCC, BamH I sites) at the 5′ of the primer, giving a total of 20 bases for each primer: 5′- GGATCCCATGNNNNNNNNNN -3′ (SEQ ID NO: 13) (Chen *et al.*, 2000). Sense primers were synthesized in 96 well format and the concentration was adjusted to 50ng/μl with TE. GLGI master mixtures were prepared for each reaction, containing 1x PCR buffer (20 mM TrisCl pH 8.4, 50 mM KCl), 2 mM $MgCl_2$, 0.2 mM dNTPs, 1.5 units / 0.3 μl PLATINUM Taq polymerase, 60 ng / 1.2 μl antisense primer (5′- ACTATCTAGAGCGGCCGCTT-3′) (SEQ ID NO: 12), and 0.5 - 5 ng of 3' cDNAs. The reaction mixtures were aliquoted into a 96-well plate at 28.8 μl per well. Sense primers (60 ng / 1.2 μl) were then added into each well. GLGI reactions were performed in PE GeneAmp PCR Systems 9600 or 9700. The conditions used were 94°C for 2 min, followed by five cycles at 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s. The conditions were then changed to 20-25 cycles at 94°C for 30 s, 60°C for 30 s, and 72°C for 30 s. Reactions were kept at 72°C for 5 min for the last cycle. The amplified products were directly precipitated in the 96-well PCR plate by adding 100μl of precipitation mixture to each well, containing 1μl of glycogen (20 mg/ml,

Roche), 15μl of 7.5M NH_4OAc and 84μl of 100% ethanol. The plate was sealed with Tape pads (QIAGEN, Inc), vortexed, and kept at room temperature for 15 min. After spinning at 4000 rpm for 35 min at 4°C (SORVALL RC5C plus; rotor: SH3000), the supernatants were removed, 150μl of 70% ethanol were added per well to wash the DNA, and the plate were spun at 4000 rpm for 15 minutes. The supernatants were removed again, the pallets were air-dried, and dissolved in 5μl of dH_2O. Two μl of DNA, 0.7 μl of salt solution, 0.7 μl of water, and 6 ng of pCR4-TOPO vector were used for each ligation reaction with TOPO TA cloning kit for sequencing (Invitrogen). The ligation reactions were performed at room temperature for 25 min. For transformation, 2 μl of ligation were mixed with 50 μl of TOPO10 competent cells (Invitrogen), kept on ice for 20 min, then heated at 42°C for 30 s, and moved on ice. SOC media (250 μl) were added per well. Plate was sealed, shaken at 37°C for 60 min at 225 rpm. The transformants were spread on LB plates containing 50 ng/ml of kanamycin and grew over night at 37°C. Positive clones were screened by direct colony-PCR. PCR master mixtures were prepared, containing 1x PCR buffer (10 mM TrisCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$), 0.1 mM dNTPs, 0.5 units / 0.1 μl Taq polymerase (TaKaRa), 60 ng of sense primer (M13 reverse primer) and 60 ng of antisense primer (M13 forward (-20) primer). The reaction mixtures were aliquoted into a 96-well plate at 25μl per well, and colonies were picked into the reaction mixtures with sterile pipette tips. PCR was performed in PE GeneAmp PCR Systems 9600 or 9700. The conditions used were 94°C for 2 min, followed by 25 cycles at 94°C for 30 s, 55°C for 30 s, and 72°C for 60 s. The reactions were kept at 72°C for 5 min after the last cycle. 75μl of precipitation mixture were added per well to precipitate DNAs, containing 22 μl of dH_2O, 15μl of 2M $NaClO_4$ and 38 μl of 2-propanol. The plate was sealed, vortexed, and kept at room temperature for 5 min. After spinning at 4000 rpm for 35 min at 4°C, the supernatants were removed, 150μl of 70% ethanol were added per well to wash the DNA, and the plate were spun at 4000 rpm for 25 minutes. Supernatants were removed again, the pallets were air-dried, and dissolved in 10μl of dH_2O. Sequencing mixtures were prepared in a total volume of 7μl , containing 0.8μl of big-dye pre-mixture, 1.4μl of dilution buffer (400 mM TrisCl pH 9.0, 10 mM $MgCl_2$), 30 ng / 0.3 μl of sequence primer (M13 reverse primer or M13 forward (-20) primer), 1.5μl H2O, and 3μl of DNA templates. Sequencing reactions were performed at 96°C for 10 s, 50°C for 5 s, and 60°C for 4 min for 99 cycles. The final sequencing products were precipitated by adding 75μl of precipitation mixture, consisting of 64μl of 100% ethanol/3M

NaOAc mixture (25:1), 1μl of glycogen (20 mg/ml) and 10μl dH_2O. The plate was sealed, vortexed, and kept at room temperature for 15 min. After spinning at 4000 rpm for 35 min at 4°C, the supernatants were removed, 150μl of 70% ethanol were added per well to wash the DNA, and the plate were spun at 4000 rpm for 15 minutes. The supernatants were removed, the pallets were air-dried, and dissolved in 3μl of loading dye. One μl was loaded in 5% sequencing gels. Four to six clones were sequenced for higher abundant SAGE tags, and 8 to 12 clones were sequenced for low abundant SAGE tags. Sequences were collected with an ABI 377 sequencer.

At page 51, line 17, to page 52, line 8, please substitute the following recitation for the existing text:

The high-throughput GLGI procedure has several differences as compared to the GLGI, for example, (i) 3′ cDNAs instead of full-length cDNAs are used as the templates for GLGI amplification. This prevents artificial amplification from non-specific annealing of sense primer to sequences upstream of the last CATG. The 3' cDNAs can be amplified to provide sufficient templates for GLGI amplification; (ii) a single antisense primer (5′-ACTATCTAGAGCGGCCGCTT-3′) (SEQ ID NO: 12) is used for all GLGI reactions instead of using combination of the five anchored oligo dT primers. The sequence of the antisense primer is located in 3′ end of all the cDNA templates incorporated from anchored oligo dT primers used for the first strand cDNA synthesis. The inventors have observed that the anchored oligo dT primers are unstable which can hinder the successful performance of GLGI. Use of the single primer also increased the efficiency of GLGI amplification significantly as any annealing of this primer with 3' end sequence results in extension during PCR. In contrast, the use of five anchored oligo dT primers results in an extension by PCR only when correctly paired primers anneal. This feature is particularly useful to amplify the templates with low copies; (iii) PLATINUM Taq polymerase instead of Pfu DNA polymerase was used for GLGI amplification, in order to increase the yield of final products, while maintaining high specificity; (iv) the GLGI amplified DNAs were directly precipitated and cloned into vector without gel purification, to prevent the loss of amplified products. This is contemplated be particularly important for products with short sizes and for products generated from templates with low copies. The inventors data showed that these changes significantly increase efficiency and specificity for GLGI amplification of 3′ cDNAs, especially for templates expressed at low level.